

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:36 ; Search time 170.72 Seconds
(without alignments)
17.102 Million cell updates/sec

Title: US-09-331-631a-5_COPY_33_75
Perfect score: 248
Sequence: 1 NQEDPQTECCQCCRRCKRQEE.....RQOYCCRRCKEICEEEY 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR66:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	44.4	566	2 S22477	vicilin precursor
2	105	42.3	588	1 FMCNAB	alpha-globulin B p
3	104	41.9	509	2 S08059	alpha-globulin typ
4	96	38.7	411	2 T29475	hypothetical prote
5	94	37.9	605	2 S06398	alpha-globulin typ
6	78.5	31.7	425	2 T18592	hypothetical prote
7	78.5	31.7	600	2 T18593	hypothetical prote
8	77	31.0	572	2 T29880	hypothetical prote
9	77	31.0	1513	2 T23681	hypothetical prote
10	75.5	30.4	242	2 T29699	hypothetical prote
11	75	30.2	388	2 T31887	hypothetical prote
12	75	30.2	388	2 T31888	hypothetical prote
13	74	29.8	438	2 T31889	hypothetical prote
14	74	29.8	445	2 T31898	hypothetical prote
15	73.5	29.6	810	2 T44430	protein pvi100 [imp
16	73	29.4	330	2 T25169	hypothetical prote
17	72.5	29.2	335	2 T31559	hypothetical prote
18	71.5	28.8	335	2 T31560	hypothetical prote
19	71.5	28.8	335	2 T31561	hypothetical prote
20	71	28.6	47	2 JCS557	arginine/glutamate
21	68.5	27.6	356	2 S31574	hypothetical prote
22	68	27.4	154	2 T27967	hypothetical prote
23	67.5	27.2	314	2 T27886	hypothetical prote
24	66.5	26.8	273	2 T20930	hypothetical prote
25	66.5	26.8	637	2 S35221	globulin Bg1 prec
26	66.5	26.8	654	2 T30136	hypothetical prote
27	66	26.6	111	2 T20452	hypothetical prote
28	65.5	26.4	33	2 A41822	antimicrobial pept
29	65	26.2	419	2 T18450	hypothetical prote

30	65	26.2	910	2 A34721	androgen receptor
31	65	26.2	911	2 B34721	androgen receptor
32	65	26.2	919	2 A39248	androgen receptor
33	65	26.2	1306	2 T13592	hypothetical prote
34	64	25.8	648	1 T01150	protein kinase (EC
35	63.5	25.6	51	1 HSMSS1	protamine - mouse
36	63.5	25.6	51	2 S03997	protamine 1 - rat
37	63	25.4	1390	2 T14004	histidine kinase h
38	62	25.0	1969	2 T08875	histidine kinase h
39	62	25.0	4957	2 T03455	ALR protein - huma
40	62	25.0	5262	2 T03454	ALR protein - huma
41	61	24.6	223	1 KASHS2	alpha-s2-casein pr
42	61	24.6	429	2 S29565	apolipoprotein A-I
43	60	24.2	285	1 I46207	involucrin - dog
44	60	24.2	1038	2 T02634	rep protein homolo
45	60	24.2	1094	2 S49313	protein kinase - s

ALIGNMENTS

```

RESULT 1
S22477
vicilin precursor - cacao
C:Species: Theobroma cacao (cacao)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: S22477; S22478; S18105; S22050
R:McHenry, L.; Fritz, P.J.
Plant Mol. Biol. 18, 1173-1176, 1992
A:Title: Comparison of the structure and nucleotide sequence of vicilin genes of coco
A:Reference number: S22477; MID:92288309
A:Accession: S22477
A:Molecule type: DNA
A:Residues: 1-566 <MCH>
A:Cross-references: EMBL:X62625
A:Accession: S22478
A:Molecule type: mRNA
A:Residues: 1-452 <MC2>
A:Cross-references: EMBL:X62626
C:Genetics:
A:Introns: 211/1; 269/3; 296/3; 391/3; 502/1
C:Superfamily: glycinin
C:Keywords: seed; storage protein
F:1-24/Domain: signal sequence #status predicted <Sig>
F:25-566/Product: vicilin #status predicted <MAT>

Query Match 44.4%; Score 110; DB 2; Length 566;
Best Local Similarity 47.5%; Pred. No. 0.00028;
Matches 19; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

QY 2 QEDPTECCQCCRRCKRQEE...RQOYCCRRCKEICEEE 41
DB 78 EEDLRQYQCCGRCQEDPQOQCCRRCKEYKRE 117

RESULT 2
FMCNAB
alpha-globulin B precursor (clone C72) - upland cotton
W:Alternate names: seed storage protein; vicilin precursor
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C:Accession: A30838; S06911
R:Chlan, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.
Plant Mol. Biol. 7, 475-489, 1986
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII
A:Reference number: A30838
A:Accession: A30838
A:Molecule type: mRNA
A:Residues: 1-588 <CHI>
A:Cross-references: GB:M16891; NID:9167374; PIDN:AAA33071.1; PID:9167375
A:Experimental source: var. Coker 201
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.

```

Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Seed
A:Reference number: 506398
A:Accession: S06911
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-81 <CH2>
C:Comment: This is a seed storage protein.
C:Superfamily: glycinin
C:Keywords: glycoprotein; seed; storage protein
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-588/Product: alpha-globulin storage preprotein #status predicted <MAT>
E:417/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	42.3%	Score 105;	DB 1;	Length 588;
Best Local Similarity	47.5%	Pred. No. 0.00093;		
Matches 19;	Conservative 9;	Mismatches 10;	Indels 2;	Gaps 1
QY	3	EDPTECCQCCRCRQESDPPQOQYCCRCRKEICEEEE	42	
		:		
		:		
DB	81	EDPQRRYECCQDECRQEE--RQRPCCCQRCRLKRRFEQD	118	

RESULT 3
S08059
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
N:Alternate names: seed storage protein
C:Species: *Gossypium hirsutum* (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S08059
R:Chlan, C.A.; Borrolo, K.; Kamalay, J.A.; Dure III, L.
Plant MOL. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. St
A:Reference number: S06398
A:Accession: S08059
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-509 <CHL>
C:Superfamily: glycinin

Query Match	41.9%	Score 104	DB 2	Length 509
Best Local Similarity	48.7%	Pred. No. 0.0011		
Matches	19	Conservative	8	Mismatches 10
				Indels 2
				Gaps 1
QY	4	DPTQCQCQRCRCRQDESDPRQDQTCQRCRCKETCEEEEE	42	
		:::::		
Db	1	DPQRYEEQCQECRCQEE--RQSPQCQCQCLRLRFEEQQ	37	

RESULT 4
T29475
hypochemical protein T01D1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence,revision 15-Oct-1999 #text,change 21-Jan-2000
C:Accession: T29475
R:Bradshaw, H.; Wohlmann, P.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid T01D1.
A:Reference number: 220623
A:Accession: T29475
A:Status: preliminary; translated from GR/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-411

A:Cross-references: EMBL:U80455; PIDN:AA837887.1; GSPDB:GN00020; CESP:T01D1.6
A:Experimental source: strain Bristol N2; clone T01D1
C:Genetics:
A:Gene: CESP:T01D1.6
A:Map position: 2
A:Introns: 25/3; 304/3
A:Superfamily: gltadin

Query Match	38.7%	Score 96	DB 2	Length 411
Best Local Similarity	39.5%	Pred. No. 0.0059		
Matches 17	Conservative 11	Mismatches 11	Indels 4	Caps 2

```

QY      2 QEPQTETCQ-CQRCRQGESDPQQQYICRRCHEICEEEFY 43
      | : | : | | | | : | : | : | : | : | : | : |
DB     138 QQPAPQQCQCQDCQACPQQQ--QPQQQCQCQCQTTCQSDQY 17

```

RESULT 5
506398

C:Species: *Gossypium hirsutum* (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S06398
R:Chan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX
A:Reference number: S06398
A:Accession: S06398
A:Status: not compared with conceptual translation
A:Status: type: DNA
A:Residues: 1-605 <CHL>
C:Superfamily: glycinin
F1:24/Domain: signal sequence #status predicted <SIG>
F1:25-605/Product: alpha-globulin type A #status predicted <MAT>

Query Match	37.9%;	Score 94;	DB 2;	Length 605;
Best Local Similarity	35.9%;	Pred. No. 0.012;		
Matches 14;	Conservative 12;	Mismatches 13;	Indels 0;	Caps 0

```

Qy      2 QEDPQTECQCQCRCRCQESDPQQQYQCGRCKEICEEE 40
      | : | : ::|||: | | : : | : | : | : | : | :
Db    115 QQQPDQKQKEQQRCQMQEQRRERKQCCVCKEGRQYQED 15

```

RESULT 6
T18592
hypothetical protein AC3.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18592
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z18595
A:Accession: T18592
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <WTL>
A:Cross-references: EMBL:Z71177; PIDN:CAA94867.1; GSPDB:IGN00023; CESP:AC3.3
A:Experimental source: clone AC3
C:Genetics:
A:Gene: CESP:AC3.3
A:Map position: 5
A:Introns: 18/3
C:Superfamily: gliadin

Query Match	31.7%;	Score 78.5;	DB 2;	Length 425;
Best Local Similarity	27.9%;	Pred. No. 0.36;		
Matches 17;	Conservative 9;	Mismatches 12;	Indels 23;	Gaps 2

```
Qy      2 QEDPTECQCQRRCR-----QEENP-----RQQYCORCKEICE 38
       | : ||||| | : ||| | : ||| : | :
Db    255 QQPSAPQCQCQONTCCQAAPVCQDQCACPQCQQSAPACQCCTSCQQTQCCQOOCCTPCQ 314
```

81	315	:	315
Db	315	Q	315

```

RESULT      7
             718593
hypothetical protein AC3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18593
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z18995
A:Accession: T18593
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-600 <NTL>
A:Cross-references: EMBL:Z71177; PIDN:CAA94868.1; GSPDB:GN00023; CESP:AC3.4
C:Experimental source: clone AC3
C:Genetics:
A:Gene: CESP:AC3.4
A:Map position: 5
A:Introns: 292/2: 374/3: 572/2

```

Query Match	31.7%;	Score 78.5;	DB 2;	length 600;
Best Local Similarity	27.9%;	Pred. No. 0.47;		
Matches 17; Conservative	9;	Mismatches 12;	Indels 23;	Gaps 2

```

QY      2  QEDPOTEGGCGCRRCR-----GQSPDP-----RQDQYCRCKKEICE 38
Dbb     122  QQSAPQCQCQCCTTCQQAAPVCCGCCAPQCQCQDQSAFACCCGCTSCGQYRQCCQDCCCTPCQC 181

```

QY	39	E	39
		:	
Db	182	Q	182

```

RESULT      8
T29880      T29880
C:Species:  Caenorhabditis elegans
C:Date:      15-Oct-1989 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29880
R:Nhan, M.; Le, T.T.
Submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F19G12.
A:Reference number: Z20704
A:Accession: T29880
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-572 <NHA>
A:Cross-references: EMBL:U51997; PTDN:AA048159.1; GSPDB:GN00028; CESP:F19G12.7
A:Experimental source: Strain Bristol NZ; clone F19G12
C:Genetics:
A:Gene: CESP:F19G12.7
A:Map position: X
A:Introns: 18/3

```

```

Query Match      31.0%; Score 77; DB 2; Length 572;
Best Local Similarity 34.0%;
Pred. No. 0.64;
Matches 17; Conservative 8; Mismatches 7; Indels 18; Gaps 3.

OY      8  ECQCQCRRCRQ-----QESDPRQOYCQRCKE---TCEBE 40
          |||||  |||  |||  |||  |||  |||  |||  |||
Ob 381  QCQCQCQNTCCQFAPYCEECQACSLCHQBPAPCCQ--CQNTCCQFAPYCRQ 429
          |||||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 9
T23681
hypothetical protein M0269.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23681
C:Matthews, L.

Submitted to the EMBL Data Library, November 1996
A:Reference number: Z19781
A:Accession: T23681
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1513 <MIL>
A:Cross-references: EMBL:Z81573; PIDN:CAB04625.1; GSPDB:GN00020; CESP:M02G9.1
A:Experimental source: Clone M02G9
C:Genetics:
A:Gene: CESP:M02G9.1
A:Map position: 2
A:Introns: 23/3; 71/3; 121/3; 183/1; 1083/3; 1141/3; 1408/1; 1449/1

A; Introns: 23/3; 71/3; 121/3; 183/1; 1083/3; 1141/3; 1408/1; 1449/1

Query Match	31.08;	Score 77;	DB 2;	Length 1513;
Query Match	31.08;	Score 77;	DB 2;	Length 1513;

Matches 17; Conservative 2; Mismatches 16; Indels 4; Gaps 2;

```
QY      3 EDPTQECQ-CRRRCQESDPPQQQYCRRCKEICEE 40
      | | | | | : | | | | | : |
Db 1141 EQQQCPQQYTCGTGGCGHNPDR--VCQNVCEVCASE 1176
```

RESULT 10

hypothetical protein F31A3.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*

C:\Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:\Accession: T29699

R;Murray, J.; Le, T.T.

submitted to the EMBL Data Library, May 1996
A:Description: The sequence of C. elegans cosmid F31A3.

A; Reference number: Z20667

A:Accession: T29699
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecular function: DNA

A: Molecule type: DNA
A: Residues: 1-242 <MP>

A:Cross-references: EMBL:U58742; PION:AB36856.1; GSPDB:GN00028; CESP:F3IA3.1
A:Experimental source: strain Bristol N2, clone F3IA2
A:Residues: 1-242 <MUR>

A; Experimental source: strain Bristol N2; clone F3LA3

A;Gene: CESP:F31A3.1

A;Map position: X
A;Introns: 18/3; 160/3

C; Superfamily: ultra-high-sulfur keratin

Query Match	30.48; Score 75.5; DB 2; Length 242;
-------------	--------------------------------------

Best Local Similarity 33.38; Pred. NO. 0.48;

```
QY 2 QEDPOTCCQCQRRCR-----QOESDPR-QOY--CQRCKEICEEEEE 42
   | : ||||| | : || : | ||| : | : | : : :
Db 77 QQPAAPQCQCQCNTCCQSAPVCQQQCACAPCCQQQFAPACQ-QCQNSCQQTQQ 126
```

hypothetical protein C03A7.4 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T31887
 R:Greco, T.; Bradshaw, H.; Elliott, G.
 Submitted to the EMBL Data Library, July 1997
 ;:Description: The sequence of C. elegans cosmid C03A7.

A;Reference number: Z21096

A;Accession: T31887
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecular type: DNA

A: Molecule type: DNA
A: Positives: 7-398 <GPF>

A:Cross-references: EMBL:AF016451: PTDN.AAB66001 1: GSPDB:GN00023. CESP.C03A7 4
A:Residues: 1-388 <GRE>

A: Experimental source: strain Bristol N2; clone C03A7
A, C1055 references: EMBL:AF010431, PIDN:AA800001.1; G

C;Genetics:
A;Gene: CESP:C03A7.4

A;Map position: 5
A;Introns: 75/3
C;Superfamily: gliadin

Query Match	30.28;	Score 75;	DB 2;	Length 388;
Best Local Similarity	37.58;	Pred. No. 0.77;		
Matches	15;	Conservative	7;	Mismatches 16;
			Indels	2;
			Gaps	2

OY 2 QEDPTEC-QCCQRRCQDESPPR-QYYCRCKEICEE 39
 | : : ||||| : || : : : :
Dd 194 QAQCQPCCQQCCSSCVQQQQQNSQCPCACNMTCSDICQ 233

RESULT 12
T31888

C:\date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:\Accession: T31888

A;Description: The sequence of C. elegans cosmid C03A7
A;Reference number: Z21096
A;Accession number: M31000

A: Molecule type: DNA
A: Residues: 1-388 <GRE>
A: Cross-references: EMBL: AF016451. PDB: 2ABE006 1.

C:Genetics:
A:Gene: CESP:C03A7.7
A:Map position: 5

C; Superfamily: gliadin

Query Match	30.28;	Score 75;	DB 2;	Length 388;
Best Local Similarity	37.58;	Pred. No. 0.77;		
Matches	15;	Conservative	7;	Mismatches 16;
			Indels	2;
			Gaps	2

Oy 2 QEDPOTEC-QOCORCRROESDPRQ-QYCYCRCKEICEE 39
| : | | | | : | : | : | :
Db 194 GACCCPQCQQCCSCSYQQQSQNQCCEACNTGCDSDTQQ 233

RESULT 13
T31889

C:\Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:\Accession: T31889

A;Description: The sequence of *C. elegans* cosmid C03A7
A;Reference number: Z21096

```

A;Molecule type: DNA
A;Residues: 1-438 <GRE>

```

A:Gene: CESP:C03A7.8

A; Introns: 75/3
C; Superfamily: gliadin

Query Match	29.8%;	Score 74;	DB 2;	Length 438;
Best Local Similarity	30.2%;	Pred. NO. 1.1;		
Matches 13;	Conservative 12;	Mismatches 12;	Indels 6;	Gaps 2

```

07      3  EDPQTECQ-QCQRR-----KQESDPRQQYQCRRKCEICEE 39
      :  :  ||| |||:  :  ||:  :  :  :  :  :  :  :  :  :  :
db     191  QQAQAQCGQCGCGCGQSSCVCQQQPTSTQCERACNMQCSIDLQ 23

```

RESULT 14
T31898

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T31898
C:Accession: T31898

A;Description: The sequence of *C. elegans* cosmid C03A7
A;Reference number: Z21096

A: Molecule type: DNA
A: Residues: 1-445 <GRE>
A: Accession: EMBL: M016451.1. GI: 2266607.1. CDS: 200003.2. CDS: 20127.14.

C;Genetics:
A;Gene: CESP:C03A7.14
A;Map position: F

C;Superfamily: gliadin

Query Match	29.8%;	Score 74;	DB 2;	Length 445;
Best Local Similarity	30.2%;	Pred. No. 1.1;		
Matches	13;	Conservative	12;	Mismatches
			12;	Indels
				Gaps
				2

```

QY      3  EDPTTECQ-QCQARC-----HQESPPRQDQYCRKCEICEE 39
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      191  QQAQAQCQPCQCQCQCQSSCVQDQDPTQCEPACMTQCSDICQ 233

```

RESULT 15
T44430
protein pv100 [imported] - winter squash

C: Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C: Accession: T44430
R: Yamada K : Shimada T : Kondo M : Nishimura M : Hara-Nishimura I

A:Reference number: T22767; MUID:99107919
A:Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single
A:Accession: TAAA30

A:Molecule type: mRNA
A:Residues: 1-810 <YAM>
A:Cross-references: EMBL:AB019195. NTD:G3808061. PTDN:BA344056.1. PID:G3808062

Query Match	29.6%	Score 73.5	DB 2	Length 810
Best Local Similarity	34.9%	Pred. No. 1.9		
Matches 15, Conservative	9	Mismatches 18	Indels 1	Gaps 1

QY 1 NDE-DPOTECQCCQRRKQQESDPRQOYCQRRCKETCEEEEE 42
 || : : : : || : || : : : || :
 Db 69 NQGSPPAEYEVCRLRQVAENGVEQQRKCEYCEENLRERQ 111

Search completed: March 1, 2001, 15:52:37
Job time: 561 sec